Hindawi
Shock and Vibration
Volume 2017, Article ID 5913041, 11 pages
https://doi.org/10.1155/2017/5913041

Research Article

K-Means Cluster for Seismicity Partitioning and Geological Structure Interpretation, with Application to the Yongshaba Mine (China)

Xueyi Shang,[1] Xibing Li,[1] A. Morales-Esteban,[2] Longjun Dong,[1] and Kang Peng[3,4,5]

[1] *School of Resources and Safety Engineering, Central South University, Changsha, China*
[2] *Department of Building Structures and Geotechnical Engineering, University of Seville, Seville, Spain*
[3] *State Key Laboratory of Coal Mine Disaster Dynamics and Control, Chongqing University, Chongqing, China*
[4] *State Key Laboratory of Resources and Safe Mining, CUMT, Xuzhou, China*
[5] *College of Resources and Environmental Science, Chongqing University, Chongqing, China*

Correspondence should be addressed to Kang Peng; pengkang@cqu.edu.cn

Received 20 December 2016; Accepted 15 June 2017; Published 30 July 2017

Academic Editor: Mariano Artés

Seismicity partitioning is an important step in geological structure interpretation and seismic hazard assessment. In this paper, seismic event location (X, Y, Z) and Euclidean distance were selected as the K-Means cluster, the Gaussian mixture model (GMM), and the self-organizing maps (SOM) input features and cluster determination measurement, respectively, and 1516 seismic events ($M > -1.5$) obtained from the Yongshaba mine (China) were chosen for the cluster analysis. In addition, a Silhouette and Krzanowski-Lai- (KL-) combined S-KL index was proposed to obtain the possible optimum cluster number and to compare the cluster methods. Results show that the K-Means cluster obtains the best cluster "quality" with higher S-KL indexes on the whole and meaningful clusters. Furthermore, the optimal number for detailed geological structure interpretation is confirmed as eleven clusters, and we found that two areas probably have faults or caves, and two faults may be falsely inferred by mine geologists. Seismic hazard assessment shows that C5 and C7 ($K = 11$) have a high mean moment magnitude (m_M) and C1, C2, C3, and C4 ($K = 11$) have a relatively high m_M, where special attention is needed when mining. In addition, C7 ($K = 11$) is the most shear-related area with a mean S-wave to P-wave energy ratio ($m_{Es/Ep}$) of 41.21. In conclusion, the K-Means cluster provides an effective way for mine seismicity partitioning, geological structure interpretation, and seismic hazard assessment.

1. Introduction

Microseismic monitoring has been widely used in mining [1–3], where seismicity partitioning is an important step in geological structure interpretation and seismic hazard assessment [4]. Although this task can be implemented by an expert visually, the expert knowledge-based cluster is non-quantitative, subjective, and hard to interpret for a large amount of data [5]. Intelligent cluster analysis, taking advantages of spatial, temporal, and other features of seismic events, is more highlighted due to its advances in extracting useful knowledge and finding hidden information from numerous seismic data [4]. Motivated by these facts, various cluster methods have been introduced/proposed for seismicity partitioning, among which K-Means cluster is the most popular one. However, it is still hard to determine the K-Means cluster number, and few works have been done to compare it with other cluster methods.

The general outline of this paper is shown in Figure 1. Firstly, the K-Means cluster, taking advantages of seismic event location features (X, Y, Z), was applied to cluster 1516 seismic events ($M > -1.5$) obtained from the Yongshaba mine (China). Then, to test the effectiveness of the K-Means cluster, the Gaussian mixture model (GMM) and the self-organizing maps (SOM) were selected as comparisons. In addition, the Silhouette and Krzanowski-Lai- (KL-)

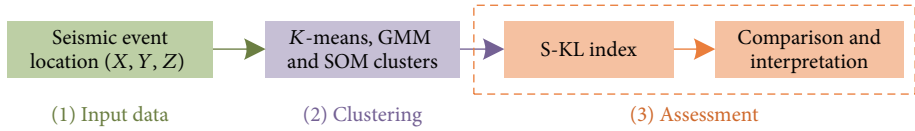


Figure 1: General outline of this paper.

combined S-KL index was proposed to obtain possible optimum cluster numbers and to compare the cluster methods. Results show that the K-Means cluster obtains a best cluster "quality" with higher S-KL indexes on the whole and continuous seismic clustering areas. The optimal number of clusters for detailed geological structure interpretation is confirmed as eleven clusters, and seismic hazard assessment shows that C5 and C7 ($K = 11$) have a high mean moment magnitude (m_M), and C1, C2, C3, and C4 ($K = 11$) have a relatively high m_M, where special attention is needed when mining. In addition, C7 ($K = 11$) is the most shear-related area with a mean S-wave to P-wave energy ratio ($m_{Es/Ep}$) of 41.21, and we found that two areas probably have faults or caves, and two faults may be falsely inferred by mine geologists.

2. State of the Art

The commonly used seismicity partitioning methods include the K-Means cluster, the hierarchical cluster, the self-organizing maps (SOM), the fuzzy cluster, the Gaussian mixture model (GMM), the density-based clustering algorithm (DBSCAN), and some other cluster means, which have been listed in Table 1.

Hierarchical cluster has been applied for seismicity partitioning since the 1990s, and Davis and Frohlich [11], Frohlich and Davis [12], Wardlaw et al. [13], Hudyma [14], Hudyma and Potvin [15], Hashemi and Mehdizadeh [16], and Hashemi and Karimi [17] selected the hierarchical clustering technique using single-link analysis/Ward's method to evaluate spatial and temporal properties of earthquake catalogues. However, the hierarchical cluster has a tendency to form seismic events into linear groups [12]. SOM is another widely used seismic cluster method, and Zamani and Hashemi [5] introduced a self-organized tectonic zoning for Iran, and then Zamani et al. [18] chose Wilk's Lambda criterion and a relative discrepancy of Wilk's Lambda to determine the optimum tectonic zoning number; and Mojarab et al. [19] discussed the effect of SOM input parameters. Fuzzy clusters have also been used in seismic clusters, Ansari et al. [4] and Benitez et al. [21] used the Gath and Giva (GG) fuzzy cluster for Iran and South West Colombia, respectively; while Monem and Hashemy [20] applied the fuzzy c-means (FCM) and the Gustafson-Kessel (GK) clusters to the Ghazvin canal irrigation network. In addition, the density-based algorithms [22–24] and the TriGen-based method [25] have also been applied for the seismic cluster. A comparison of using different seismic clusters can be found in Leśniak and Isakow [26] and Konstantaras et al. [27].

K-Means cluster analysis can be applied to the hypocentral distribution of observed earthquakes in any region [6], which has been widely used in seismic cluster. Weatherill and Burton [6] implemented the K-Means cluster to provide a catalogue of earthquakes for Aegea, and results demonstrate that from 20 to 30 clusters are the most appropriate number in capturing the hypocentral distribution and fault type. Rehman et al. [7] applied the K-Means cluster to estimate seismic hazard and risk in Pakistan and found that 19 clusters were the best to include the sufficient knowledge of subsurface geological structure and geophysical information. Morales-Esteban et al. [8] proposed an adaptive Mahalanobis K-Means algorithm to study the seismic catalogues of the Croatia and Iberian Peninsula, and the result showed the ability of this method to discover not only circular but also elliptical shapes. Ramdani et al. [9] took advantages of earthquake centroids of the K-Means cluster to find subduction evidence beneath the Gibraltar Arc and Andean regions. Besheli et al. [10] used the K-Means cluster to partition Iran into six main clusters and to study the importance of foreshock precursors for different zones.

It can be seen that most seismic clusters are focused on earthquake data, and few works have been done on mine seismic clusters. Therefore, the cluster adaptively to mine seismic events should be discussed. Though the K-Means cluster has been widely used now, it is still hard to determine the K-Means cluster number and few works have been done to compare it with other cluster methods. In addition, Table 1 shows that most optimum cluster numbers are below 20, and the Silhouette index and the Krzanowski-Lai (KL) index are the most frequently used indexes to evaluate cluster "quality." In this paper, the K-Means cluster was applied to partition 1516 seismic events ($M > -1.5$) obtained from the Yongshaba mine (China) and the GMM and SOM clusters were selected as comparisons, and the cluster results with cluster numbers from 2 to 20 have been evaluated by the Silhouette and Krzanowski-Lai- (KL-) combined S-KL index.

3. Methodology

3.1. K-Means Cluster. The K-Means cluster is a hard partitioning algorithm proposed by Hartigan and Wong [28]. A dataset of n seismic events with p dimensions consists of an $n \times p$ matrix (1), and the K-Means partitions the dataset into K clusters with each element allocated to a particular cluster.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad (1)$$

where each row represents one seismic event and each column represents one variable.

Table 1: Summary of typical seismicity partitioning methods.

No.	Cluster method	Study area	Index for possible optimum cluster number selection	Optimum cluster number	Reference
1	*K*-Means cluster	Aegean region	Krzanowski–Lai index	20~30	[6]
2	*K*-Means cluster	Pakistan	Krzanowski–Lai index	19	[7]
3	*K*-Means cluster	Croatia and Iberian Peninsula	Simplified Silhouette Width Criterion, Davies–Bouldin index, Calinski–Harabasz index and Area index	7 and 16, respectively	[8]
4	*K*-Means cluster	Gibraltar Arc and Andean regions	Silhouette index	3~4	[9]
5	*K*-Means cluster	Iran	Silhouette index	6	[10]
6	Hierarchical cluster	Global dataset	Cutoff link length	—	[11–13]
7	Hierarchical cluster	A Canadian mine	Cutoff link length	10	[14, 15]
8	Hierarchical cluster	Zagros region (Iran)	Cutoff link length	5	[16]
9	Hierarchical cluster	United States	Cutoff link length	10	[17]
10	SOM cluster	Iran	Cutoff similarity level	2, 3, 4, 6 and 22	[5]
11	SOM cluster	Iran	Wilk's Lambda and its discrepancy	11	[18]
12	SOM cluster	Iran	Dunn, Davies-Bouldin, Silhouette, C, Calinski–Harabasz, Hartigan and Krzanowski–Lai indexes	9~13	[19]
13	Fuzzy cluster	Iran	Partition density index	16	[4]
14	Fuzzy cluster	Ghazvin canal	Separation index, Xie-Beni index and partition index	12	[20]
15	Fuzzy cluster	South West Colombia	Fuzzy hyper volume and partition density	12	[21]
16	DBSCAN cluster	Japan	Correlation Ecorr and minimum number of objects in a cluster	10	[22]
17	Seismic mass DBSCAN cluster	Hellenic Arc	Correlation Ecorr and minimum number of objects in a cluster	35~40	[23]
18	Point density based cluster	Burmese–Andaman and West Sunda Arc	Cutoff link length	13	[24]
19	TriGen based cluster	Iberian Peninsula	Fitness function	34	[25]
20	MMD, Hierarchical and *K*-means clusters	Zabrze-Bielszowice coal mine (Poland)	Minimum number of events	6	[26]
21	FCM, DBSCAN, QC and dynamic spatial cluster	Hellenic Arc	—	19	[27]

The K-Means cluster is an iterative process with K initial cluster centroids, and each datum is allocated to its nearest cluster centroid. Then, the mean of each group is calculated to obtain a new cluster centroid. The procedure terminates when no datum changes clusters or the iteration number reaches a preset maximum. The details of the K-Means cluster are as follows.

Step 1. Import dataset $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, set the maximum iteration number, and randomly generate K points to be initial cluster centroids $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_K$.

Step 2. Calculate the Euclidean distance between each datum and each cluster centroid, and the data is allocated to its nearest cluster centroid.

Step 3. Calculate the mean of each cluster by $\mathbf{m}_k = \sum_{i=1}^{\|C_k\|} \mathbf{x}_i / \|C_k\|$, where $\mathbf{x}_i \in C_k$ and $\|C_k\|$ is the data number in cluster C_k, $k = 1, 2, \ldots, K$.

Step 4. Has any datum changed its cluster?

(1) Yes: continue and repeat Steps 2–5.

(2) No: output cluster each datum allocates to and list the cluster centroids.

Step 5. Has it reached the maximum iteration number?

(1) No: continue, and repeat Steps 2–5.

(2) Yes: output cluster each datum allocates to and list the cluster centroids.

3.2. Cluster Validity Indexes. Assessment of the "quality" of a partition is an important consideration in cluster analysis. From Section 2, it is known that the Silhouette and the KL indexes are the most widely used, and their brief introductions are given as follows.

3.2.1. Silhouette Index. The Silhouette provides a measure of separation quality between clusters, and the higher the Silhouette index, the better the cluster results [29]. For an event i belonging to cluster C_i, the Silhouette width is defined as

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \tag{2}$$

where a_i is the average distance between event i and other events in cluster C_i and b_i is the minimum distance between event i and events in cluster $\mathbf{X} - C_i$.

The average Silhouette width S_i and the Silhouette index are defined by (3) and (4), respectively,

$$S_i = \frac{1}{\|C_i\|} \cdot \sum_{i=1}^{\|C_i\|} s_i \tag{3}$$

$$\text{Silhouette} = \frac{1}{K}\sum_{i=1}^{K} S_i = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{1}{\|C_i\|} \cdot \sum_{i=1}^{\|C_i\|} s_i\right). \tag{4}$$

3.2.2. Krzanowski-Lai Index. The KL index produces the traces of the within-clusters matrices for two consecutive clusters, and the higher the KL index, the better the cluster results [30]. The KL index is defined as

$$\text{KL}(K) = \left|\frac{\text{DIFF}(K)}{\text{DIFF}(K+1)}\right|, \tag{5}$$

where $\text{DIFF}(K) = (K-1)^{2/d} WK_{K-1} - K^{2/d} WK_K$ and d is the number of features in the dataset. WK is the sum of all point-to-point distances of observations within each cluster, summed across all clusters.

4. Application of K-Means Cluster

4.1. Data Set. We used 1516 seismic events as the cluster dataset, which were obtained from the Yongshaba mine (26°38′N, 106°37′E) from 01/01/2014 to 04/30/2014, with moment magnitude larger than −1.5 (Note: the Institute of Mine Seismology (IMS) system of the Yongshaba mine can monitor a seismic event whose moment magnitude is as small as −3; however, events too small usually have a large location error due to unclear P arrivals and less triggered sensors. In this paper, we chose a moment magnitude larger than −1.5 for seismic cluster).

The distribution of geological structures and sensors of Yongshaba mine is shown in Figure 2(a). It shows that the IMS system contains 28 sensors, including two triaxial sensors (T1 and T2) and twenty-six uniaxial sensors (sensors from 1 to 26), and the sensors have a sampling frequency of 6000 Hz. There are seven main faults, namely, F331, F310, F316, F313, F350, F302, and F309, in which the F350 is inferred by mine geologists. In addition, there are 18 small faults, which are marked from I to XVIII, and the faults V, VI, X, and XI are inferred by mine geologists. The symbols ① and ② represent two caves.

Figure 2(b) shows the two view distributions (YX and YZ views) of the seismic dataset. The color and size of the circle represent seismic moment magnitude. It can be seen that the main seismic moment magnitude is 0.0~0.5, then 0.5~1.0, and −0.5~0.0. There are some seismic moment magnitudes smaller than −0.5 (It can be interpreted as that there are many small moment magnitude seismic events not recorded by the system, though the system can monitor a seismic moment magnitude as small as −3) and few seismic moment magnitudes are larger than 1.0.

It is easy to see that the main faults are distributed in the north, then in the south, and fewer faults in the middle, which are in a good agreement with the distribution of seismic events. Therefore, we believe that there should be a good relationship between seismic events and geological structures which needs further study.

4.2. Possible Cluster Number Selection. To compare the cluster "quality" and select possible cluster numbers, the most commonly used indexes Silhouette and KL are selected, and their values of K-Means, GMM, and SOM clusters are shown in Figure 3, in which the seismic event location (X, Y, Z) is selected as cluster input variables.

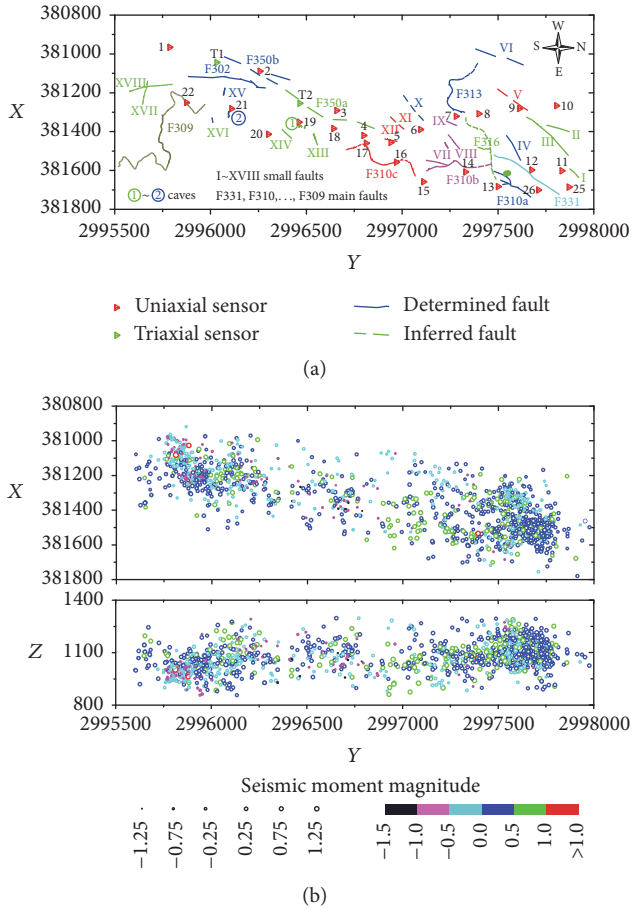


FIGURE 2: Distributions of geological structures, sensors, and seismic events of the Yongshaba mine. (a) The distributions of geological structures and sensors. The solid lines represent determined faults, while the dashed lines represent inferred faults by mine geologists. The symbols ① and ② represent two caves, and the triangles represent the location of sensors. The numbers near the sensors are the sensor numbers, T1 and T2 are triaxial sensors, and the sensors from 1 to 26 are uniaxial ones. (b) The distributions of seismic events. The color and the size of the circle represent the seismic moment magnitude.

It is easy to note that the Silhouette values within 4 clusters are higher than those of cluster numbers larger than 4 and that the K-Means cluster has a higher Silhouette index than that of the GMM and the SOM clusters on the whole. The three clusters all have a high KL value when $K = 2$, and there are some local maximum KL values. Therefore, it is hard to determine possible cluster numbers by considering both Silhouette and KL indexes. To solve this problem, a Silhouette and KL combined S-KL index (6) is proposed, which takes advantage of the product of the normalized Silhouette index and KL index, and the normalization is calculated by (7). The relatively large S-KL values correspond to possible optimum cluster numbers. Generally, the K-Means cluster has higher S-KL indexes than those of the GMM and the SOM clusters, which means that the K-Means obtains a best cluster "quality" among the three clusters, and the cluster numbers 2, 3, 6, 9, 11 and 15 are selected as possible cluster numbers for further study.

$$\text{S-KL} = \text{Silhouette}_{\text{normalization}} \cdot \text{KL}_{\text{normalization}} \tag{6}$$

$$x_{\text{normalization}} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}. \tag{7}$$

4.3. K-Means Cluster Results. Seismicity partitioning with few clusters produces a simple and general classification defining the most basic structure of the area under investigation; however, it is hard to interpret a detailed geological structure, while for a too large cluster number, there may be some clusters that are hard to interpret [7]. Therefore, a balanced cluster number should be selected as the optimum cluster number to interpret the detailed geological structure. Five typical variables, number of seismic events (N), maximum moment magnitude ($M_{\max}$), mean moment magnitude (m_M), mean S-wave to P-wave energy ratio ($m_{\text{Es/Ep}}$), and number of events with moment magnitude >0.5 ($N_{0.5}$), are selected for seismic hazard assessment and seismic source mechanism analysis, where $m_{\text{Es/Ep}}$ is a measure to evaluate the seismic source mechanism, and the higher the $m_{\text{Es/Ep}}$ is, the more the geological structure is shear-related [14].

The cluster results and typical variables of 2, 3, 6, 9, 11, and 15 clusters are shown in Figure 4, and the clustering process diagram and geological structure interpretation of 11 clusters are shown in Figure 5 for better understanding. Figures 4 and 5 show that the K-Means has a stable cluster result (there are some similar zones for different cluster numbers), and the number of events in each cluster has no significant difference, which indicates that the K-Means has a good cluster performance.

When the cluster number is less than 6, the cluster results are in a good agreement with regional faults (the cluster results match with known geological structures). In the two-class division (Figure 4(a)), the cluster divides the dataset into north zone (C1, concentrated faults) and south zone (C2, less faults), and C1 has many more $N_{0.5}$ events and a higher m_M than that of the C2; that is to say that the north zone is more active and dangerous than the south zone. Then the dataset was divided into north zone (C1), middle zone (C2), and south zone (C3), looking for a relation between the clustering and the existing geology. C2 has a nearly equal $N_{0.5}$ to C3, though there are much more seismic events in C3, so C2 should also be paid attention when mining. In addition, C2 is more shear-related than C3. For the six-partition division (Figure 4(c)), the clusters C1, C2, and C3 with concentrated faults have a higher m_M and more $N_{0.5}$ than C4, C5, and C6, and C5 and C6 are less shear-related.

When the cluster number is larger than 9, the cluster results mainly correspond to the detailed geological structures (most clusters for $K = 9$ just have few faults except the cluster C4). For the nine-partition division (Figure 4(d)), C2, C3, and C4 have relatively large $m_{\text{Es/Ep}}$ and m_M, while C1 and C5 have a relatively large $m_{\text{Es/Ep}}$ or m_M. C6 has a relatively large $m_{\text{Es/Ep}}$, while C7 and C9 have both low $m_{\text{Es/Ep}}$ and m_M. However, there still are many faults in C4, and a more detailed cluster should be studied.

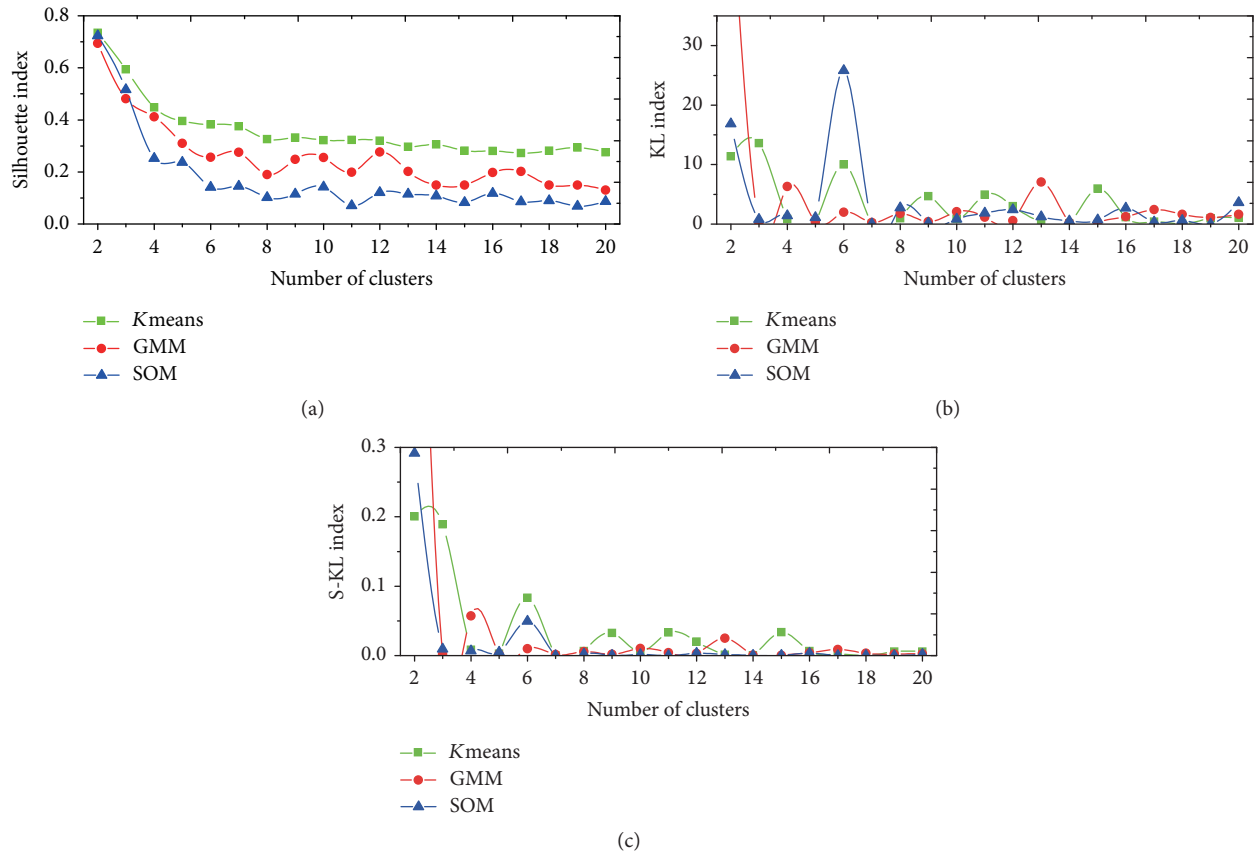


FIGURE 3: Cluster indexes for cluster "quality" comparison and possible cluster number selection of the K-Means cluster, the GMM cluster, and the SOM cluster. (a) Silhouette index; (b) KL index; (c) S-KL index.

The 11-partition division (Figure 4(d)) separates C4 (K = 9) into clusters C5, C6, and C7; a detailed geological structure interpretation is shown in Figure 5. It can be seen that C7 has the highest $m_{\mathrm{Es/Ep}}$ and m_M, with values of 0.33 and 41.21, respectively; thus special attention should be paid to the C7 area. Further study shows that there should be faults or caves in areas a and b, and the cluster analysis also confirms that there should be faults X~XII, F350a, and F350b. However, there are probably no faults for VI and the bottom F309, which needs further study, while for K = 15, the clusters C3, C4, C5, and C6 are so connected, which makes it hard for geological structure interpretation, and the clusters C14 and C15 contain areas too large to find detailed geological structure information. In conclusion, the authors believe that the 11 clusters are the best for a detailed geological structure interpretation.

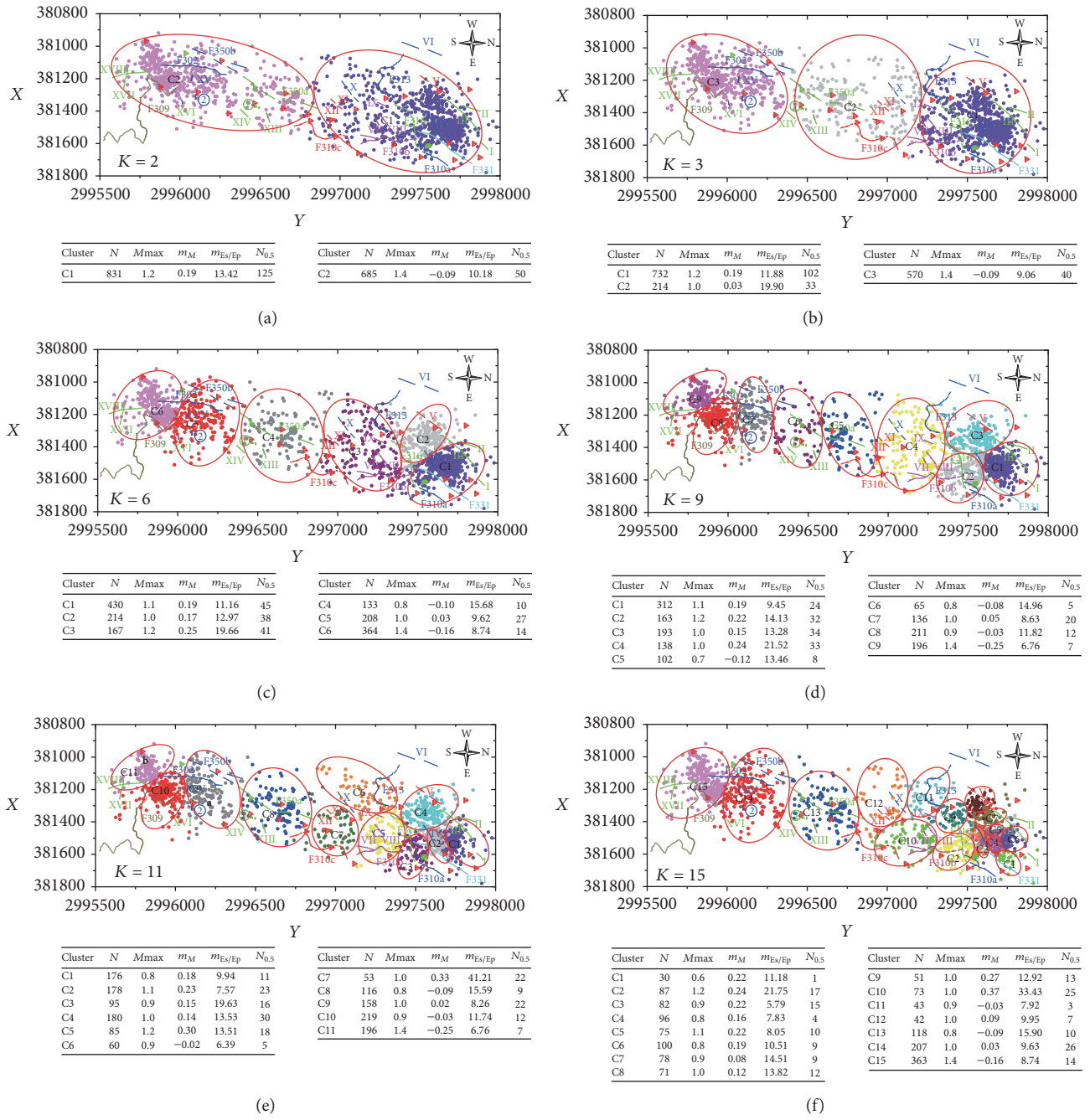
5. Cluster Comparisons

For a better comparison of the K-Means, GMM, and SOM clusters, the same cluster number of K = 11 is selected and a possible optimum cluster number near 11 is also discussed in this analysis (K = 13 for the GMM cluster and K = 12 for the SOM cluster), and the cluster results are shown in Figure 6.

It can be observed that there are some discontinuous zones for both GMM and SOM clusters. For the GMM cluster: (1) K = 11, C1 surrounds C2 and C3, and C6 is isolated, while C8 and C9 are mixed together; (2) K = 13, C1 also surrounds C2 and C3, and C8 is isolated. For K = 11 and K = 12 of the SOM cluster, the clusters C1, C2, C3, and C4 are mixed together, while C7 is isolated. In addition, there are small seismic event number clusters for both GMM and SOM clusters. For example, C6 and C11 (K = 11) for the GMM cluster have small seismic event numbers. In conclusion, the results of the GMM and SOM clusters cannot be considered satisfactory, and they partition some areas with no meaningful information.

6. Discussions

The above analysis has shown that the K-Means cluster is effective in seismicity partitioning for geological structure interpretation. The GMM cluster is on the basis that the K clusters follow the Gauss distributions, and its aim is to make the probability function maximum. The GMM cluster can be affected by the seed number, and the GMM cluster results (K = 11) of seed = 10, 25, 50, and 100 are shown in Figure 7(b). It is easy to observe that all the four GMM cluster results have discontinuous zones or cross-section, such as the C1, C2, C9, and C10 when seed number is equal to 10. This can be interpreted as that the neighboring seismic event locations sometimes are not in the Gauss distribution and



(a)

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C1	831	1.2	0.19	13.42	125

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C2	685	1.4	−0.09	10.18	50

(b)

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C1	732	1.2	0.19	11.88	102
C2	214	1.0	0.03	19.90	33

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C3	570	1.4	−0.09	9.06	40

(c)

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C1	430	1.1	0.19	11.16	45
C2	214	1.0	0.17	12.97	38
C3	167	1.2	0.25	19.66	41

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C4	133	0.8	−0.10	15.68	10
C5	208	1.0	0.03	9.62	27
C6	364	1.4	−0.16	8.74	14

(d)

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C1	312	1.1	0.19	9.45	24
C2	163	1.2	0.22	14.13	32
C3	193	1.0	0.15	13.28	34
C4	138	1.0	0.24	21.52	33
C5	102	0.7	−0.12	13.46	8

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C6	65	0.8	−0.08	14.96	5
C7	136	1.0	0.05	8.63	20
C8	211	0.9	−0.03	11.82	12
C9	196	1.4	−0.25	6.76	7

(e)

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C1	176	0.8	0.18	9.94	11
C2	178	1.1	0.23	7.57	23
C3	95	0.9	0.15	19.63	16
C4	180	1.0	0.14	13.53	30
C5	85	1.2	0.30	13.51	18
C6	60	0.9	−0.02	6.39	5

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C7	53	1.0	0.33	41.21	22
C8	116	0.8	−0.09	15.59	9
C9	158	1.0	0.02	8.26	22
C10	219	0.9	−0.03	11.74	12
C11	196	1.4	−0.25	6.76	7

(f)

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C1	30	0.6	0.22	11.18	1
C2	87	1.2	0.24	21.75	17
C3	82	0.9	0.22	5.79	15
C4	96	0.8	0.16	7.83	4
C5	75	1.1	0.22	8.05	10
C6	100	0.8	0.19	10.51	9
C7	78	0.9	0.08	14.51	9
C8	71	1.0	0.12	13.82	12

Cluster	N	Mmax	m_M	$m_{Es/Ep}$	$N_{0.5}$
C9	51	1.0	0.27	12.92	13
C10	73	1.0	0.37	33.43	25
C11	43	0.9	−0.03	7.92	3
C12	42	1.0	0.09	9.95	7
C13	118	0.8	−0.09	15.90	10
C14	207	1.0	0.03	9.63	26
C15	363	1.4	−0.16	8.74	14

FIGURE 4: Cluster results and typical variables of K-Means with different cluster numbers. (a) $K = 2$; (b) $K = 3$; (c) $K = 6$; (d) $K = 9$; (e) $K = 11$; (f) $K = 15$. The solid lines represent determined faults, while the dashed lines represent inferred faults by mine geologists. The symbols ① and ② represent two caves, and the triangles represent the location of the sensors. The statistics under (a)–(f) are the typical variables for each cluster, where N is the number of seismic events, M_{max} is the maximum moment magnitude, m_M is the mean moment magnitude, $m_{Es/Ep}$ is the mean S-wave to P-wave energy ratio, and $N_{0.5}$ is the number of events with moment magnitude >0.5.

the close seismic events may not be subjected to the same Gauss distribution. The SOM converts high dimension input space into a low dimension representation (typically two dimension), which has a very good image of visualizing. The SOM cluster can be affected by the learning rate, ordering Epochs, and lattice width and height. The lattice width × height is set to 1 × 11; then we randomly selected ten combinations of learning rate (0.1, 0.3, 0.5, 0.7, and 0.9) and ordering Epochs (2000, 3000, 5000, and 10000) to test the SOM cluster ($K = 11$). However, we obtain the same result with Figure 6(c); this may be due to the fact that the input variable just has a few dimensions (three dimensions), which makes the SOM cluster find an optimum result with different parameters. However, research has shown that the SOM may

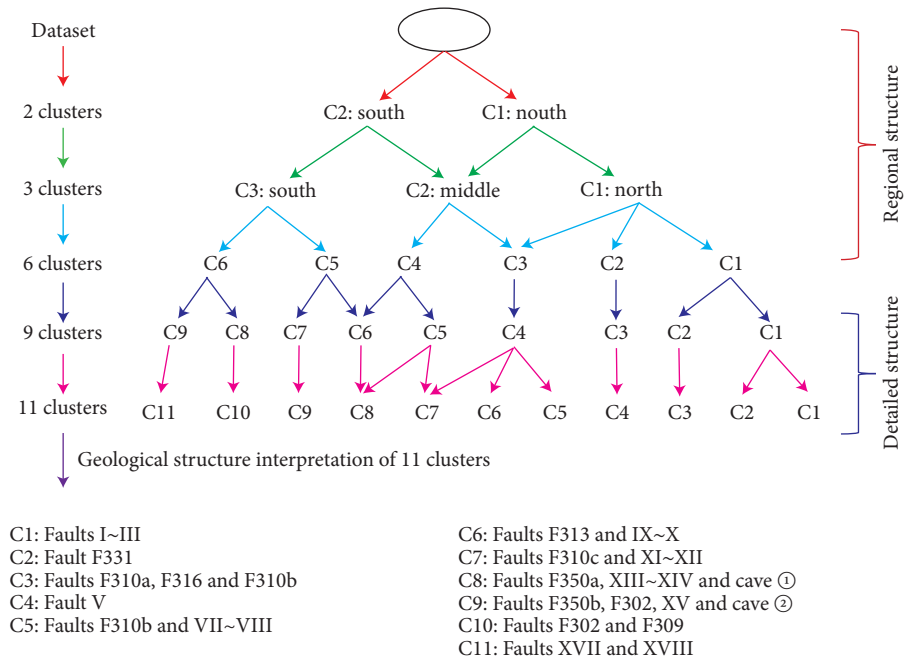


Figure 5: K-Means clustering process diagram and geological structure interpretation of 11 clusters for Figure 4.

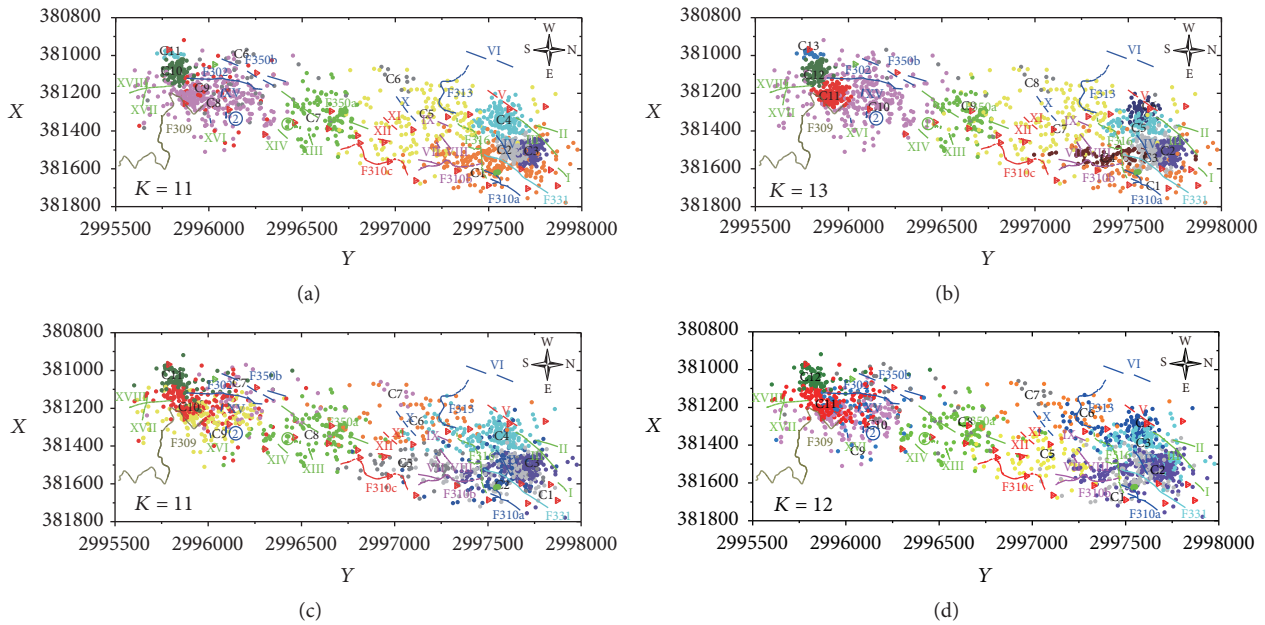


Figure 6: Cluster results of GMM and SOM with different cluster numbers. (a) and (b): $K = 11$ and $K = 13$ of GMM clusters; (c) and (d): $K = 11$ and $K = 12$ of SOM clusters. The solid lines represent determined faults, while the dashed lines represent inferred faults by mine geologists. The symbols ① and ② represent caves, and the triangles represent the location of sensors.

only be valid for zoning high seismic activity areas [25], and, for low seismic activity zones, it may have bad cluster results. The K-Means cluster can be affected by the initial clustering centers, and the typical K-Means cluster results ($K = 11$) are shown in Figure 7(a). It is easy to see that most K-Means cluster zones are the same, though there may be some difference using different initial clustering centers: there may be some seismic events that are not in the same cluster, for example, C7 and C8 between Figures 7(a1) and 7(a2); the clusters may be divided into different clusters or combined into one cluster, for example, C1 and C2 in Figure 7(a1) are equal to C1 in Figure 7(a3), and C10 in Figure 7(a1) is equal to C6 and C10 in Figure 7(a3). However, there is no assumption of Gauss distribution for the K-Means cluster, and it partitions closer events to one cluster. This makes the K-Means cluster obtain continuous cluster zones and help us

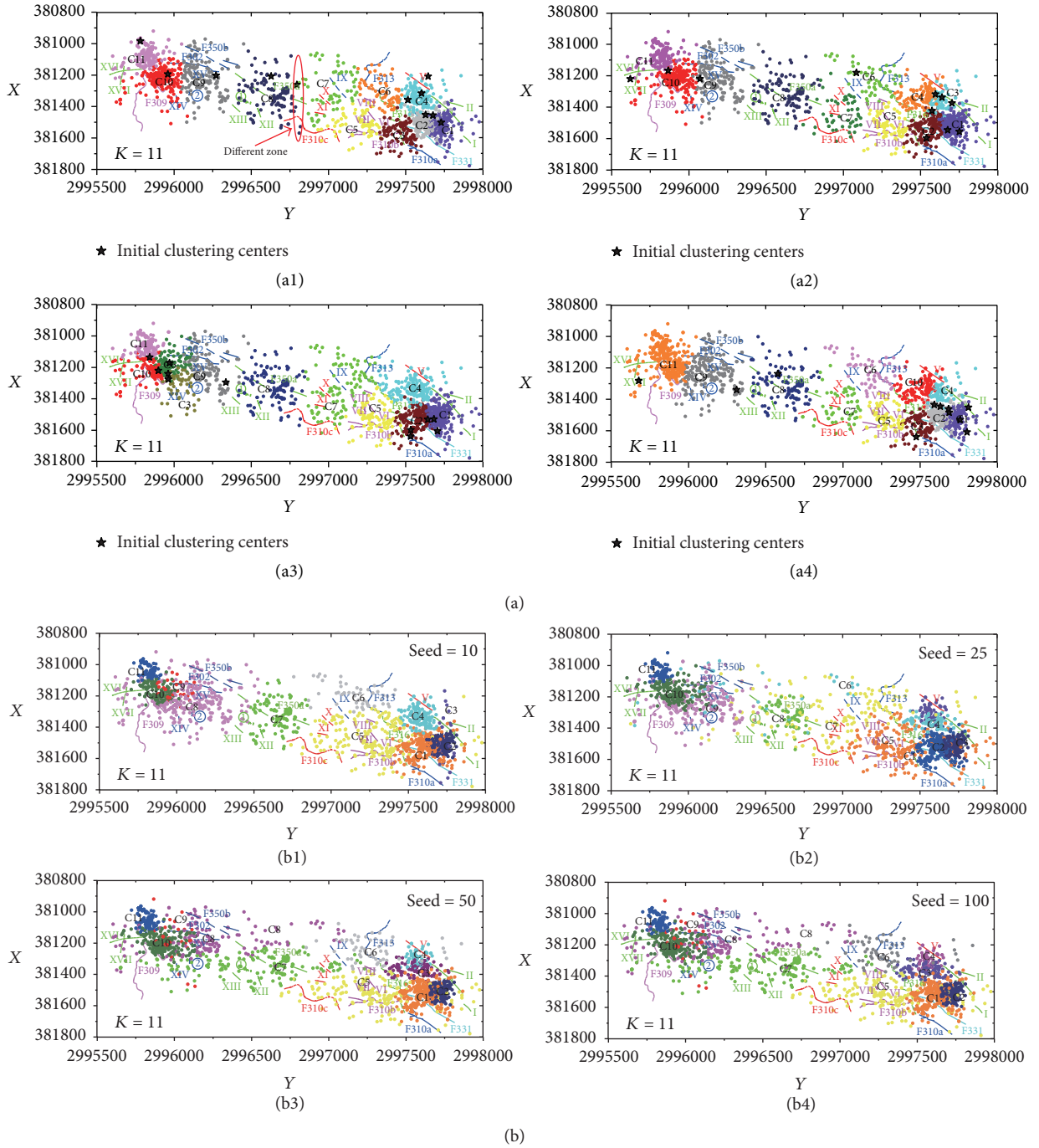


FIGURE 7: Cluster results ($K = 11$) of K-Means and GMM with different cluster parameters. (a) K-Means cluster with different initial clustering centers; (b) GMM cluster with different cluster seeds. (b1) Seed = 10; (b2) Seed = 25; (b3) Seed = 50; (b4) Seed = 100. The solid lines represent determined faults, while the dashed lines represent inferred faults by mine geologists. The symbols ① and ② represent two caves, and the triangles represent the location of sensors.

better interpret the relation between seismicity and geological structure.

7. Conclusions

Three intelligent spatial clusters, the K-Means cluster, the GMM, and the SOM, have been applied to partition the seismicity of the Yongshaba mine (China). Then, the Silhouette and KL combined S-KL index was applied to obtain possible optimum cluster numbers and to compare the cluster methods. Results show that the K-Means cluster obtains the best cluster "quality" with higher S-KL indexes on the whole and meaningful clusters. The optimal cluster number for detailed geological structure interpretation is confirmed to

be eleven clusters, and we found that two areas probably have faults or caves, and two faults may be falsely inferred by mine geologists. Furthermore, seismic hazard assessment shows that C5 and C7 ($K = 11$) have a high mean moment magnitude (m_M), and C1, C2, C3, and C4 ($K = 11$) have a relatively high m_M, where special attention is needed when mining. In addition, C7 ($K = 11$) is the most shear-related area, with a mean S-wave to P-wave energy ratio ($m_{Es/Ep}$) of 41.21. In conclusion, the K-Means cluster provides an effective way for mine seismicity partitioning, geological structure interpretation, and seismic hazard assessment.

Additional Points

Highlights. (i) The K-Means cluster has been applied for seismicity partitioning, geological structure interpretation, and seismic hazard assessment. (ii) The possible optimum cluster number is determined by the proposed Silhouette and Krzanowski-Lai- (KL-) combined S-KL index. (iii) The K-Means cluster has a better seismicity partitioning "quality" for geological structure interpretation than that of the Gaussian mixture model (GMM) and the self-organizing maps (SOM).

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the financial support of the National Key Research and Development Program of China (2016YFC0600706), National Natural Science Foundation of China (51504044), the Research Fund of The State Key Laboratory of Coal Resources and Mine safety Project (14KF05), CUMT, China, and the Research Fund of Chongqing Basic Science and Cutting-Edge Technology Special Projects (cstc2016jcyjA1861).

References

[1] L. J. Dong, X. B. Li, and G. N. Xie, "Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive Bayes classification," *Abstract and Applied Analysis*, Article ID 459137, Art. ID 459137, 8 pages, 2014.

[2] L. J. Dong, J. Wesseloo, Y. Potvin, and X. B. Li, "Discrimination of Mine Seismic Events and Blasts Using the Fisher Classifier, Naive Bayesian Classifier and Logistic Regression," *Rock Mechanics and Rock Engineering*, vol. 49, no. 1, pp. 183–211, 2016.

[3] L. J. Dong, J. Wesseloo, Y. Potvin, and X. B. Li, "Discriminant models of blasts and seismic events in mine seismology," *International Journal of Rock Mechanics & Mining Sciences*, vol. 86, pp. 282–291, 2016.

[4] A. Ansari, A. Noorzad, and H. Zafarani, "Clustering analysis of the seismic catalog of Iran," *Computers & Geosciences*, vol. 35, no. 3, pp. 475–486, 2009.

[5] A. Zamani and N. Hashemi, "Computer-based self-organized tectonic zoning: a tentative pattern recognition for Iran," *Computers & Geosciences*, vol. 30, no. 7, pp. 705–718, 2004.

[6] G. Weatherill and P. W. Burton, "Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region," *Geophysical Journal International*, vol. 176, no. 2, pp. 565–588, 2009.

[7] K. Rehman, P. W. Burton, and G. A. Weatherill, "K-means cluster analysis and seismicity partitioning for Pakistan," *Journal of Seismology*, vol. 18, no. 3, pp. 401–419, 2014.

[8] A. Morales-Esteban, F. Martinez-Alvarez, S. Scitovski, and R. Scitovski, "A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning," *Computers & Geosciences*, vol. 73, pp. 132–141, 2014.

[9] F. Ramdani, O. Kettani, and B. Tadili, "Evidence for subduction beneath Gibraltar Arc and Andean regions from k-means earthquake centroids," *Journal of Seismology*, vol. 19, no. 1, pp. 41–53, 2015.

[10] P. R. Besheli, M. Zare, R. Ramezani Umali, and G. Nakhaeezadeh, "Zoning Iran based on earthquake precursor importance and introducing a main zone using a data-mining process," *Natural Hazards*, vol. 78, no. 2, pp. 821–835, 2015.

[11] S. D. Davis and C. Frohlich, "Single-Link Cluster Analysis, Synthetic Earthquake Catalogues, and Aftershock Identification," *Geophysical Journal International*, vol. 104, no. 2, pp. 289–306, 1991.

[12] C. Frohlich and S. D. Davis, "Single-Link Cluster Analysis As A Method to Evaluate Spatial and Temporal Properties of Earthquake Catalogues," *Geophysical Journal International*, vol. 100, no. 1, pp. 19–32, 1990.

[13] R. L. Wardlaw, C. Frohlich, and S. D. Davis, "Evaluation of precursory seismic quiescence in sixteen subduction zones using single-link cluster analysis," *Pure and Applied Geophysics*, vol. 134, no. 1, pp. 57–78, 1990.

[14] M. R. Hudyma, *Analysis And Interpretation of Clusters of Seismic Events in Mines*, University of Western Australia, Perth, Australia, 2009.

[15] M. R. Hudyma and Y. H. Potvin, "An engineering approach to seismic risk management in hardrock mines," *Rock Mechanics and Rock Engineering*, vol. 43, no. 6, pp. 891–906, 2010.

[16] S. N. Hashemi and R. Mehdizadeh, "Application of hierarchical clustering technique for numerical tectonic regionalization of the Zagros region (Iran)," *Earth Science Informatics*, vol. 8, no. 2, pp. 367–380, 2014.

[17] M. Hashemi and H. A. Karimi, "Seismic source modeling by clustering earthquakes and predicting earthquake magnitudes," in *Smart City 360°*, vol. 166, Springer International Publishing, Cham, Switzerland, 2016.

[18] A. Zamani, M. Khalili, and A. Gerami, "Computer-based self-organized tectonic zoning revisited: Scientific criterion for determining the optimum number of zones," *Tectonophysics*, vol. 510, no. 1-2, pp. 207–216, 2011.

[19] M. Mojarab, H. Memarian, M. Zare, A. H. Morshedy, and M. H. Pishahang, "Modeling of the seismotectonic provinces of Iran using the self-organizing map algorithm," *Computers & Geosciences*, vol. 67, pp. 150–162, 2014.

[20] M. J. Monem and S. M. Hashemy, "Extracting physical homogeneous regions out of irrigation networks using fuzzy clustering method: A case study for the Ghazvin canal irrigation network," *Journal of Hydroinformatics*, vol. 13, no. 4, pp. 652–660, 2011.

[21] H. D. Benitez, J. F. Florez, D. P. Duque, A. Benavides, O. L. Baquero, and J. Quintero, "Spatial pattern recognition of seismic events in South West Colombia," *Computers & Geosciences*, vol. 59, pp. 60–77, 2013.

[22] S. J. Nanda and G. Panda, "Design of computationally efficient density-based clustering algorithms," *Data & Knowledge Engineering*, vol. 95, pp. 23–38, 2015.

[23] G. Georgoulas, A. Konstantaras, E. Katsifarakis, C. D. Stylios, E. Maravelakis, and G. J. Vachtsevanos, "'Seismic-mass' density-based algorithm for spatio-temporal clustering," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4183–4189, 2013.

[24] B. Mukhopadhyay, M. Fnais, M. Mukhopadhyay, and S. Dasgupta, "Seismic cluster analysis for the Burmese-Andaman and West Sunda Arc: Insight into subduction kinematics and seismic potentiality," *Geomatics, Natural Hazards and Risk*, vol. 1, no. 4, pp. 283–314, 2010.

[25] F. Martínez-Álvarez, D. Gutiérrez-Avilés, A. Morales-Esteban, J. Reyes, J. L. Amaro-Mellado, and C. Rubio-Escudero, "A novel method for seismogenic zoning based on triclustering: Application to the Iberian Peninsula," *Entropy*, vol. 17, no. 7, pp. 5000–5021, 2015.

[26] A. Leśniak and Z. Isakow, "Space-time clustering of seismic events and hazard assessment in the Zabrze-Bielszowice coal mine, Poland," *International Journal of Rock Mechanics and Mining Sciences*, vol. 46, no. 5, pp. 918–928, 2009.

[27] A. J. Konstantaras, E. Katsifarakis, E. Maravelakis, E. Skounakis, E. Kokkinos, and E. Karapidakis, "Intelligent Spatial-Clustering of Seismicity in the Vicinity of the Hellenic Seismic Arc," *Earth Science Research*, vol. 1, no. 2, 2012.

[28] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: a k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[29] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[30] W. J. Krzanowski and Y. T. Lai, "A criterion for determining the number of groups in a data set using sum-of-squares clustering," *Biometrics. Journal of the Biometric Society*, vol. 44, no. 1, pp. 23–34, 1988.